

# Insect survey 2020, basic analysis

*Jens Å*

*11 November, 2020*

## Contents

<b>Data import and cleaning</b>	<b>2</b>
Write a species list for the entomologists . . . . .	4
<b>Filter out the Nerlandsøya samples</b>	<b>5</b>
Check the emptying dates. . . . .	7
<b>Check missing weight data</b>	<b>9</b>
<b>Basic species richness graphs</b>	<b>10</b>
<b>Basic biomass graphs</b>	<b>12</b>
<b>Check overlap between malaise and window traps</b>	<b>15</b>
Total number of coleoptera species per trap type . . . . .	17
Percentage of catches in each trap type per species . . . . .	17
<b>Figs to Niclas on Pollinators</b>	<b>21</b>
<b>Phenology graphs</b>	<b>24</b>
<b>Plot some weather data</b>	<b>28</b>
<b>Sampling time analysis</b>	<b>32</b>
A similar comparison, but for biomass. . . . .	43
<b>Compare the ethanol to the propglykol window traps</b>	<b>48</b>
<b>To-do</b> * Need to compare vaatvekt minus bottles. Or better, tørrvekt.	
• These species are not matched to the gyldig names in artsdatabasen yet.	
• Should store dates and times separately as well in logger_data table. Easier to subset later.	
• Update weight figures with bottle weight removed.	
• Notera i rapport hur många treffer som er misstenkte.	
• Make plots of phenology	

- No. species over the season for each location. Split into separate figs for the two habitats?
  - Relate this to local temperature, temperature sums?
- Plots of number of samples over time
- Check differences btw 2 week and 4 week sampling
  - Species occurrences,
  - compare 2 \* 2 weeks vs. 1 \* 4 week
  - Compare first and second of the 2 weeks, see if any of the have more overlap with 4 week. If 4 week becomes full, the first 2 week period should overlap more.
  - Do this for window traps as well.
- Check overlap/additions of window traps vs. malaise traps
  - Subset only Coleoptera
  - Group by location + time
- Within habitat, check relationship btw species richness and a few variables
  - Forest age
  - Diversity of fauna
  - Temperature
  - Altitude
- Check differences btw ethanol and propglyk window traps

## Data import and cleaning

At this point I get the raw data from Marie, which requires some cleaning. In the future, we'll put this all in a database, and separate out the cleaning and database import in a separate script.

```
#dat <- read_delim("../Data/Genetikkdata/long_format_data_GCF-2020-735_736_19102020.txt",
#  as.tibble()

#dat <- read_delim("../Data/Genetikkdata/long_format_data_GCF-2020-735_736_19102020.txt"

dat <- read_delim("../Data/Genetikkdata/long_format_data_GCF-2020-735_736_744_26102020.txt"
  delim = ";",
  col_types = cols(
  kingdom = col_character(),
  phylum = col_character(),
  class = col_character(),
  order = col_character(),
  family = col_character(),
```

```

genus = col_character(),
species = col_character(),
sample = col_character(),
value = col_double(),
GenlabID = col_character(),
GUID = col_character(),
Tilsatt_antall_C.analis_SI_USA_08.06.20 = col_double(),
Tilsatt_antall_C.chinensis_08.06.20 = col_double(),
Tilsatt_antall_Sirisser = col_double(),
Tilsatt_antall_Melormer = col_double(),
Tilsatt_antall_C.maculatus = col_double(),
Vekt_tom_flaske_.g. = col_double(),
Vekt_flaske_med_EtOH_og_insekter_.g. = col_double(),
Vekt_flaske_.insekter_vaat_.g. = col_double(),
Vekt_flaske_.insekter_toerr_.g. = col_double(),
Totalvolum_.mengde_ATL_.prot.K_mL. = col_character(),
Vaatvekt_.g. = col_double(),
Toerrvekt_.g. = col_double(),
Kommentar_proeve = col_character(),
Ekstraksjonsdato = col_date(format = "%d.%m.%Y"),
Ekstraksjonskit = col_character(),
Subsamplet_volum_.ul. = col_double(),
Elueringsvolum = col_double(),
OD = col_double(),
X260_280 = col_double(),
X260_230 = col_double(),
Labperson = col_character(),
Ekstraksjonskommentar = col_character(),
locality = col_character(),
trap = col_character(),
date_placed = col_date(format = "%d.%m.%Y"),
time_placed = col_time(format = ""),
empty_week = col_double(),
dateemptied = col_date(format = "%d.%m.%Y"),
timeemptied = col_time(format = ""),
liquid = col_character(),
personemptied = col_character(),
sample_id = col_character(),
mottat_lab = col_character(),
comment = col_character()
),
na = c("NA",
      "N/A",
      "Maa veies"

```

```
)  
)
```

Fix some coding errors.

```
# dat %>%
#   select(genus, species, new_species) %>%
#   slice(297, 1902)
#
#
# dat %>%
#   select(genus, species, new_species) %>%
#   slice(3219, 3585, 4017, 11424, 13275, 15465, 21056, 24168, 25077, 31567, 32805, 48502, 49
#
#Ekstraksjonskommentar = gsub(pattern = "[\\\"]", "", Ekstraksjonskommentar),
dat <- dat %>%
  mutate(species_latin = species)

dat <- dat %>%
  mutate(species = gsub(" sp", "_sp", species))

dat <- dat %>%
  mutate(species = gsub("[0-9](_)(sp)", "\\\\[1]\\3", species))

dat <- dat %>%
  mutate(species = gsub("(.*)(_)(.*)", "\\3", species))

dat <- dat %>%
  mutate(locality = str_to_sentence(locality))
```

Set factor levels

```
dat <- dat %>%
  mutate(locality = factor(locality, levels = c(paste0("Semi-nat_", 1:10), paste0("Skog_", 1
trap_type = ifelse(grepl("MF", trap), "Malaise", "Vindu"),
habitat_type = ifelse(grepl("Skog", locality), "Skog", "SN_Gressmark"),
liquid = ifelse(grepl("etoh85", liquid), "Ethanol", "PropGl_Ethanol"),
weeks_sampled = ifelse(grepl("1", trap) | grepl("3", trap), 2, 4)
)
```

Write a species list for the entomologists

```
dat %>%
  select(kingdom,
```

```

    phylum,
    class,
    order,
    family,
    genus,
    species) %>%
distinct() %>%
arrange(kingdom,
    phylum,
    class,
    order,
    family,
    genus,
    species) %>%
write_excel_csv(path = "out/species_list_2020-11-02.csv")

```

## Filter out the Nerlandsøya samples

```

dat <- dat %>%
  filter(locality != "Nerlandsoeya") %>%
  droplevels()

```

Add missing bottle weights. This is based on a sample of bottle dry weight measurements made after the regular weighings.

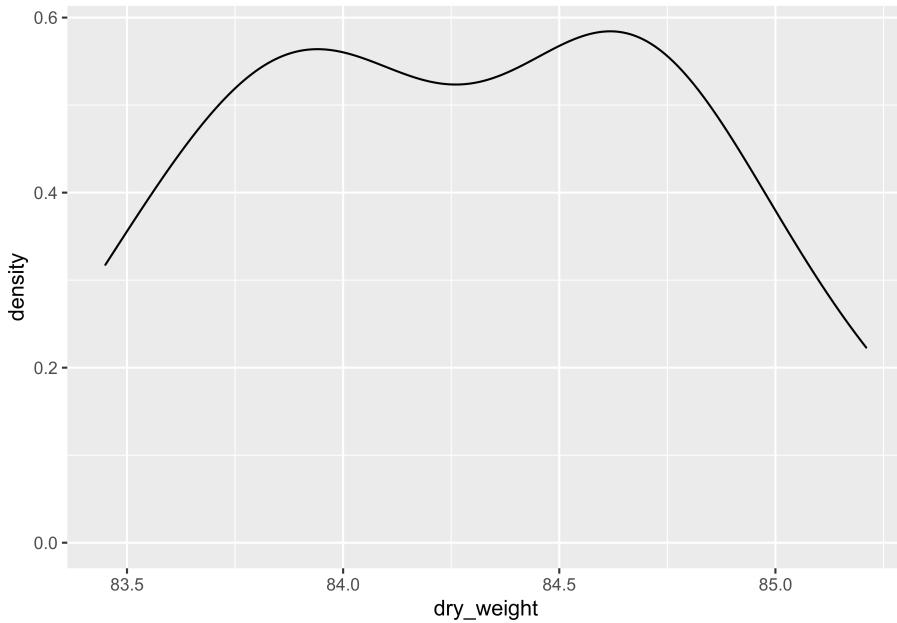
```

bottle_weights <- tbl(con,
  in_schema("catch_data", "bottle_weights")) %>%
  select(-id)

## Warning in postgresqlExecStatement(conn, statement, ...): RS-DBI driver warning:
## (unrecognized PostgreSQL field type uuid (id:2950) in column 0)

bottle_weights %>%
  filter(bottle_type == "MF" & year == 2000) %>%
  select(dry_weight) %>%
  ggplot(.) +
  geom_density(aes(x = dry_weight))

```



```

bottle_weights_MF <- bottle_weights %>%
  filter(bottle_type == "MF" & year == 2000) %>%
  select(dry_weight) %>%
  summarize(mean_weight = mean(dry_weight)) %>%
  collect()

## Warning: Missing values are always removed in SQL.
## Use `mean(x, na.rm = TRUE)` to silence this warning
## This warning is displayed only once per session.

sum(is.na(dat$Vekt_tom_flaske_.g.))

## [1] 6572

dat <- dat %>%
  mutate(Vekt_tom_flaske_.g. = ifelse(trap_type == "Malaise" & is.na(Vekt_tom_flaske_.g.), NA, Vekt_tom_flaske_.g.))

sum(is.na(dat$Vekt_tom_flaske_.g.))

## [1] 1368

dat <- dat %>%
  mutate(Vaatvekt_.g. = Vekt_flaske_..insekter_vaat_.g. - Vekt_tom_flaske_.g.,
         Toerrvekt_.g. = Vekt_flaske_..insekter_toerr_.g. - Vekt_tom_flaske_.g.)

```

## Check the emptying dates.

The dates seems to be OK this time. Should do a proper check on the endpoints though.

```
dat <- dat %>%
  mutate(days_collected = as.numeric(dat$date_emptied - dat$date_placed))

# dat %>%
#   filter(days_collected == 0) %>%
#   select(sample_id,
#         GUID,
#         date_placed,
#         date_emptied) %>%
#   distinct() %>%
#   print(n = Inf)
#
# hist(as.numeric(dat$date_emptied - dat$date_placed))

dat %>%
  group_by(days_collected) %>%
  summarise(n())

## `summarise()` ungrouping output (override with `.`groups` argument)

## # A tibble: 16 x 2
##   days_collected `n()`
##   <dbl> <int>
## 1 10     368
## 2 11    1084
## 3 12    5654
## 4 13    6223
## 5 14   10831
## 6 15    8053
## 7 16   11159
## 8 17    1422
## 9 25    561
## 10 26   1101
## 11 27   3016
## 12 28   9687
## 13 29   1132
## 14 30    450
## 15 31   1175
## 16 32    734

# dat %>%
#   filter(days_collected == 8) %>%
```

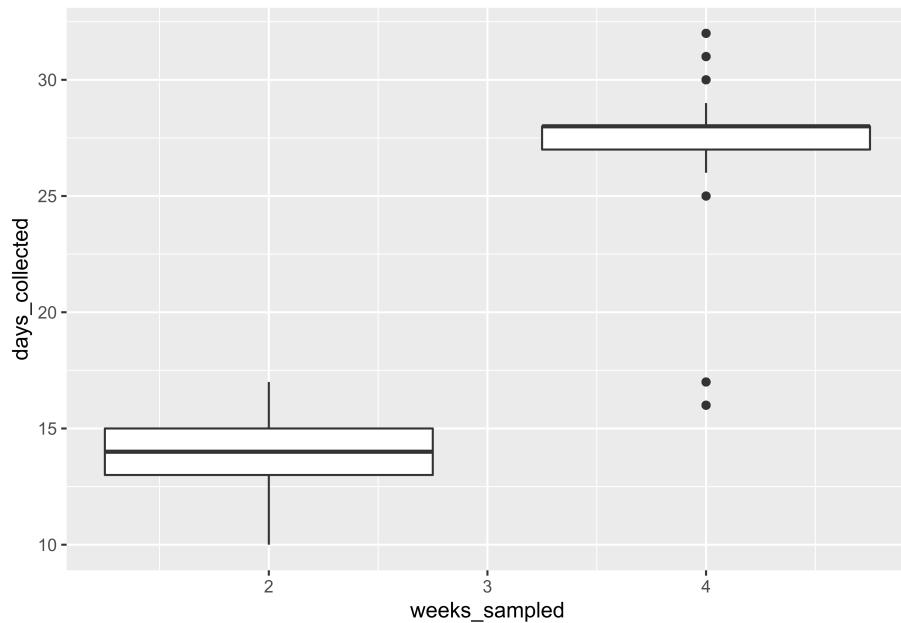
```

#   select(sample_id,
#          GUID,
#          date_placed,
#          date_emptied) %>%
#  distinct() %>%
#  print(n = Inf)

```

Set weeks sampled manually for the few locations in Oslo that got an extra short sampling in the beginning.

```
ggplot(dat, aes(x = weeks_sampled, y = days_collected, group = weeks_sampled)) +
  geom_boxplot()
```



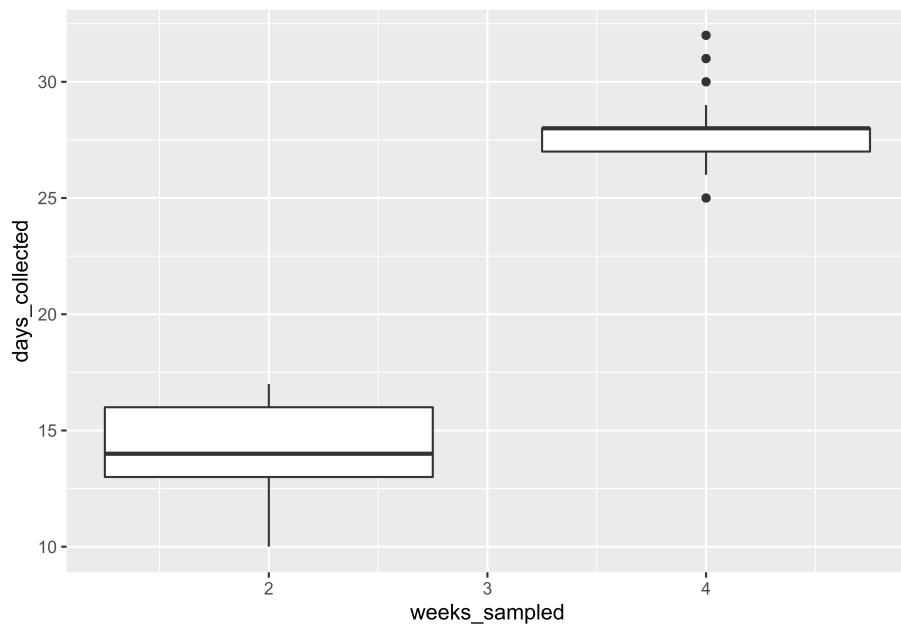
```

twoWeeks <- dat %>%
  filter(weeks_sampled == 4 &
         days_collected < 20) %>%
  select(sample_id) %>%
  distinct() %>%
  as.vector()

dat <- dat %>%
  mutate(weeks_sampled = ifelse(sample_id %in% twoWeeks[[1]], 2, weeks_sampled))

ggplot(dat, aes(x = weeks_sampled, y = days_collected, group = weeks_sampled)) +
  geom_boxplot()

```



## Check missing weight data

```

dat %>%
  group_by(empty_week) %>%
  select(Vekt_tom_flaske_.g.,
         Vekt_flaske_med_EtOH_og_insekter_.g.,
         Vekt_flaske_.insekter_vaat_.g.,
         Vekt_flaske_.insekter_toerr_.g.,
         Vaatvekt_.g.,
         Toerrvekt_.g.)
  ) %>%
  distinct() %>%
  summarise_all(~sum(is.na(.)))

```

## Adding missing grouping variables: `empty\_week`

	empty_week	Vekt_tom_flaske~	Vekt_flaske_med~	Vekt_flaske_.i~	Vekt_flaske_.i~
	<dbl>	<int>	<int>	<int>	<int>
## 1	22	6	0	0	0
## 2	24	20	0	0	0
## 3	26	10	3	0	2

```

## 4      28      4      9      0      0
## 5      30      2      2      0      2
## 6      32      0      0      0      0
## # ... with 2 more variables: Vaatvekt_.g. <int>, Toerrvekt_.g. <int>

```

## Basic species richness graphs

Quick look at diversity at the different sites.

```

total_no_taxa <- dat %>% select(order, family,genus,species) %>%
  distinct() %>%
  count()

dat %>%
  filter(grep("MF", trap)) %>%
  group_by(locality) %>%
  select(genus,species) %>%
  distinct() %>%
  count() %>%
  ungroup() %>%
  ggplot(.) +
  geom_bar(aes(y = n, x = locality, fill = locality),
            stat = "identity") +
# ggttitle(paste0("Kumulativ artsrikedom i\nalle malaisefeller per ", max(dat$dateemptied)),
#           theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1),
#                 legend.position = "none") +
#                 scale_fill_nina() +
#                 ylab("Antall taksa") +
#                 xlab("Lokalitet"))

## Adding missing grouping variables: `locality`
dat %>%
  filter(grep("VF", trap)) %>%
  group_by(locality) %>%
  select(genus,species) %>%
  distinct() %>%
  count() %>%
  ungroup() %>%
  ggplot(.) +
  geom_bar(aes(y = n, x = locality, fill = locality),
            stat = "identity") +
# ggttitle(paste0("Kumulativ artsrikedom i\nalle vindusfeller per ", max(dat$dateemptied),
#               theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1),
#                     legend.position = "none") +
#                     scale_fill_nina() +

```

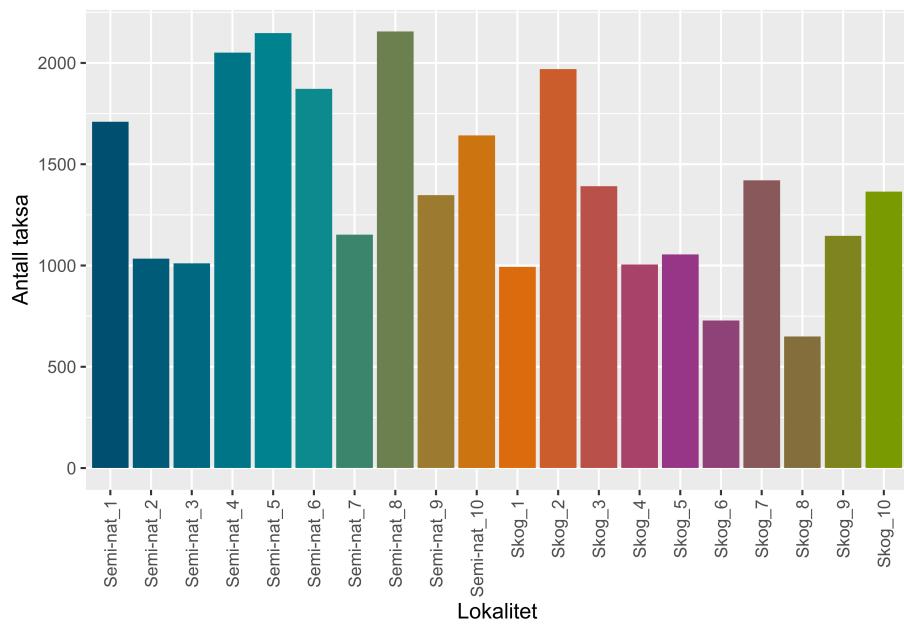


Figure 1: Cumulative number of taxa found in malaise traps.

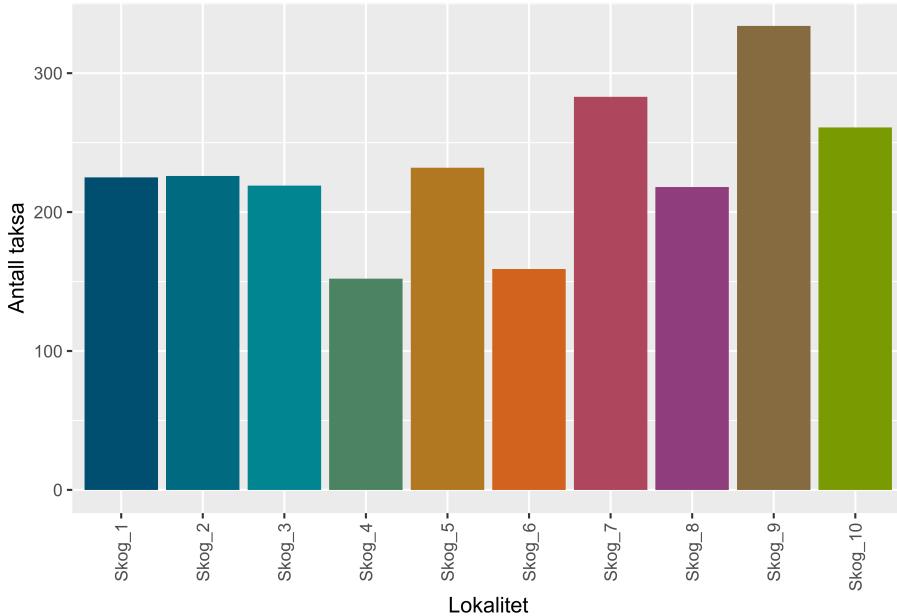


Figure 2: Cumulative number of taxa found in window traps.

```
ylab("Antall taksa") +
xlab("Lokalitet")  

## Adding missing grouping variables: `locality`
```

## Basic biomass graphs

NB, we still have some missing data on the weight of the bottles.

Quick quality check of the spread of the bottle weights. Could we impute the missing values?

```
dat %>%
  select(Vekt_tom_flaske_.g.,
         Vekt_flaske_med_EtOH_og_insekter_.g.,
         Vekt_flaske_._insekter_vaat_.g.,
         Vekt_flaske_._insekter_toerr_.g.,
         Vaatvekt_.g.,
         Toerrvekt_.g.)  

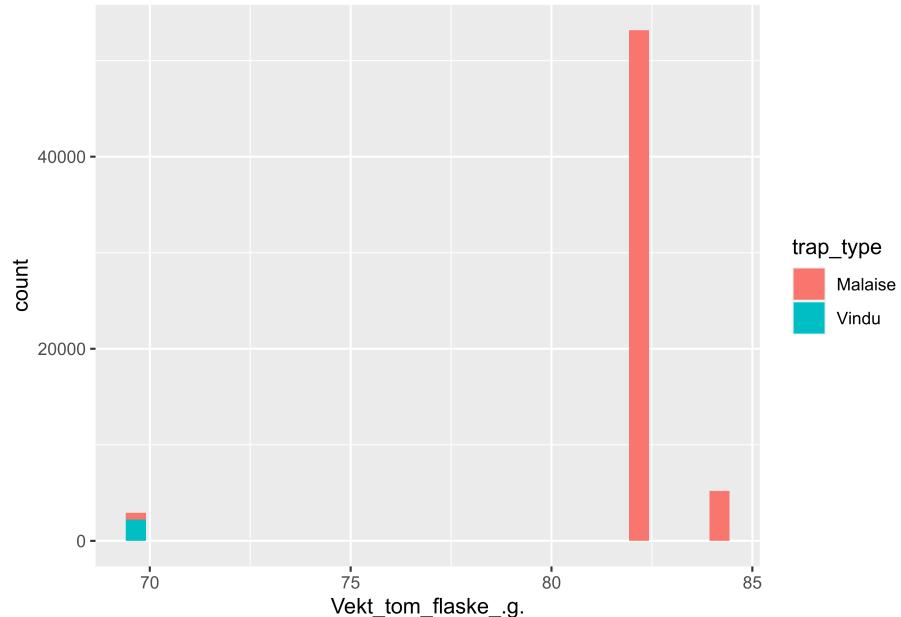
## # A tibble: 62,650 x 6
```

```

##      Vekt_tom_flaske~ Vekt_flaske_med~ Vekt_flaske_..i~ Vekt_flaske_..i~
##                <dbl>           <dbl>           <dbl>           <dbl>
## 1          82.2          728          118.          89.0
## 2          82.2          728          118.          89.0
## 3          82.2          728          118.          89.0
## 4          82.2          728          118.          89.0
## 5          82.2          728          118.          89.0
## 6          82.2          728          118.          89.0
## 7          82.2          728          118.          89.0
## 8          82.2          728          118.          89.0
## 9          82.2          728          118.          89.0
## 10         82.2          728          118.          89.0
## # ... with 62,640 more rows, and 2 more variables: Vaatvekt_.g. <dbl>,
## #   Toerrvekt_.g. <dbl>
ggplot(dat) +
  geom_histogram(aes(Vekt_tom_flaske_.g., fill = trap_type))

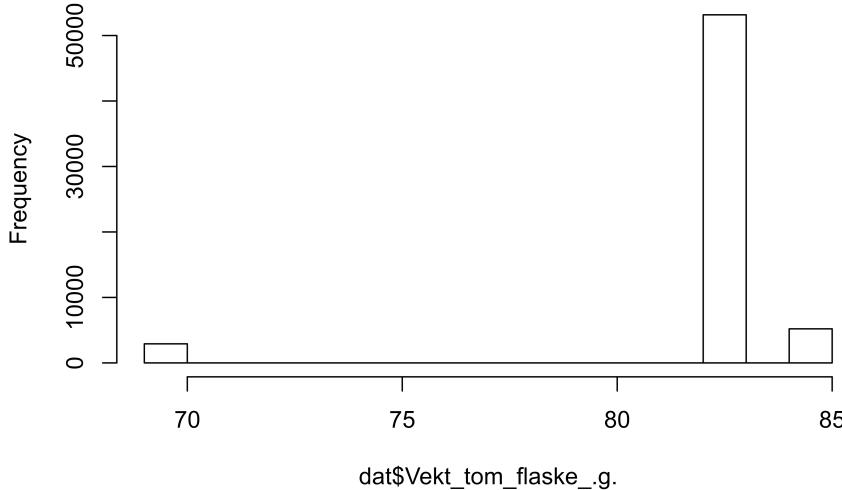
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 1368 rows containing non-finite values (stat_bin).

```



```
hist(dat$Vekt_tom_flaske_.g.)
```

Histogram of dat\$Vekt\_tom\_flaske\_.g.



```
table(dat$Vekt_tom_flaske_.g.)
```

```
##          Min. 1st Qu. Median      Mean 3rd Qu.      Max. NA's
##       69.75     82.21    82.2103 84.281666666667
##       2915     40333    12830      5204
```

These bottle weights are weirdly consistent. Did they really weigh all bottles?  
UPDATE WITH NEW BOTTLE WEIGHTS. !!DONE!!

```
summary(dat$Toerrvekt_.g.)
```

```
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max. NA's
##     -0.57     4.84    8.33    13.61   16.17    80.35    4207
```

Quick look at the biomass at the different sites.

```
dat %>%
  filter(grepl("MF", trap)) %>%
  group_by(locality) %>%
  ggplot(.) +
  geom_boxplot(aes(y = Toerrvekt_.g., x = locality, fill = locality)) +
  # ggtitle(paste0("Middels avrunnen vekt (pluss flaske)\nav insekter i malaisefeller per ",
  # theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1),
  # legend.position = "none") +
  # scale_fill_nina() +
  ylab("Vekt (g.)") +
  xlab("Lokalitet")
```

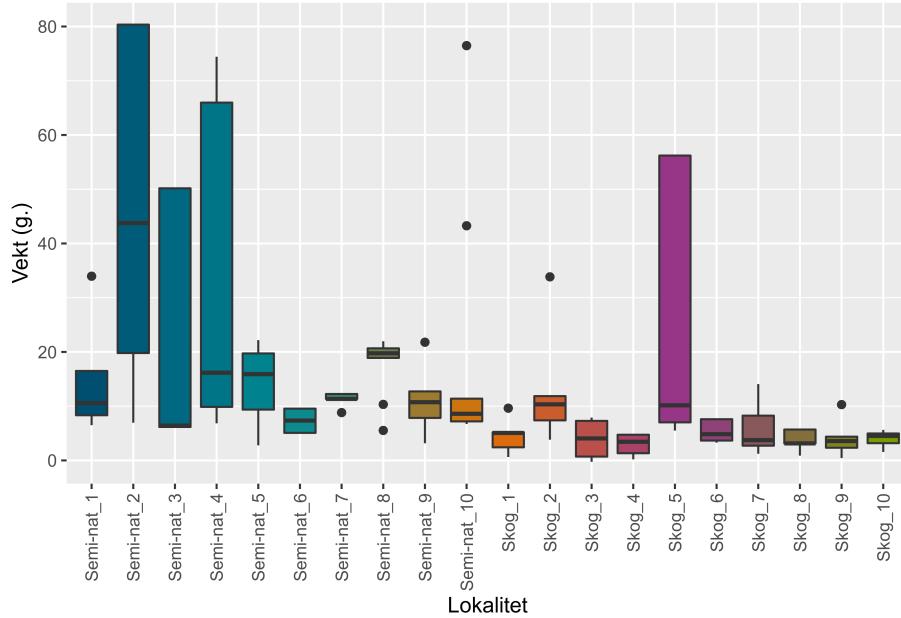


Figure 3: Average biomass (dry weight) in malaise traps.

```
## Warning: Removed 2839 rows containing non-finite values (stat_boxplot).
dat %>%
  filter(grepl("VF", trap)) %>%
  group_by(locality) %>%
  ggplot(.) +
  geom_boxplot(aes(y = Toerrvekt_.g., x = locality, fill = locality)) +
  #ggtitle(paste0("Middels avrunnen vekt (pluss flaske) av insekter i\n vindusfeller per ", 
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1),
        legend.position = "none") +
  scale_fill_nina() +
  ylab("Vekt (g.)") +
  xlab("Lokalitet")

## Warning: Removed 1368 rows containing non-finite values (stat_boxplot).
```

## Check overlap between malaise and window traps

Simple: total spec richn for all traps and times per location, plus the same for just malaise and window traps. group in figure by location

complex: tally how many spec is found in only one trap type. No unnecessary

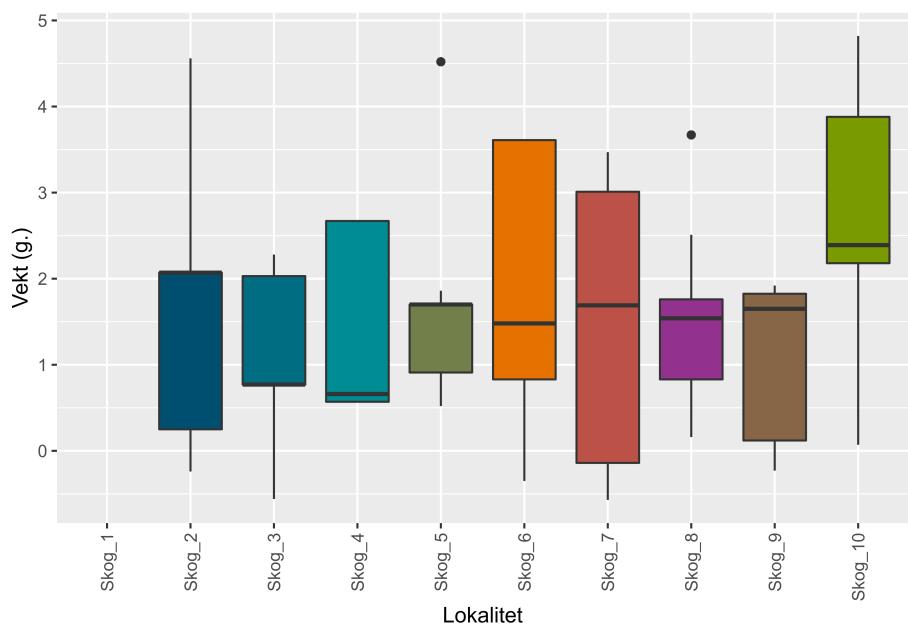


Figure 4: Average biomass (dry weight) in window traps.

per species: which sp is only found in one trap type. or per species, % of findings in each trap type

### Total number of coleoptera species per trap type

```
col_spec_richn_tot <- dat %>%
  filter(order == "Coleoptera",
         habitat_type == "Skog") %>%
  group_by(locality) %>%
  summarise(col_spec_richn = n_distinct(genus, species)) %>%
  mutate(trap_type = "Alle")

## `summarise()` ungrouping output (override with `.`groups` argument)
col_spec_richn_trap <- dat %>%
  filter(order == "Coleoptera",
         habitat_type == "Skog") %>%
  group_by(locality,
           trap_type) %>%
  summarise(col_spec_richn = n_distinct(genus, species))

## `summarise()` regrouping output by 'locality' (override with `.`groups` argument)
col_spec_richn <- col_spec_richn_tot %>%
  union_all(col_spec_richn_trap)

ggplot(col_spec_richn, aes(y = col_spec_richn,
                           x = locality,
                           fill = trap_type)) +
  geom_bar(stat = "identity",
            position = position_dodge()) +
#  ggtitle(paste0("Kumulativt antall billearter (identifiserte taksa)\\ni hver felletyp per"))
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  scale_fill_nina(name = "Felletype") +
  ylab("Antall taksa") +
  xlab("Lokalitet")
```

### Percentage of catches in each trap type per species

How to make it more readable? Split out all the singletons into a separate graph?

```
col_perc_traptype <- dat %>%
  filter(order == "Coleoptera",
         habitat_type == "Skog") %>%
```

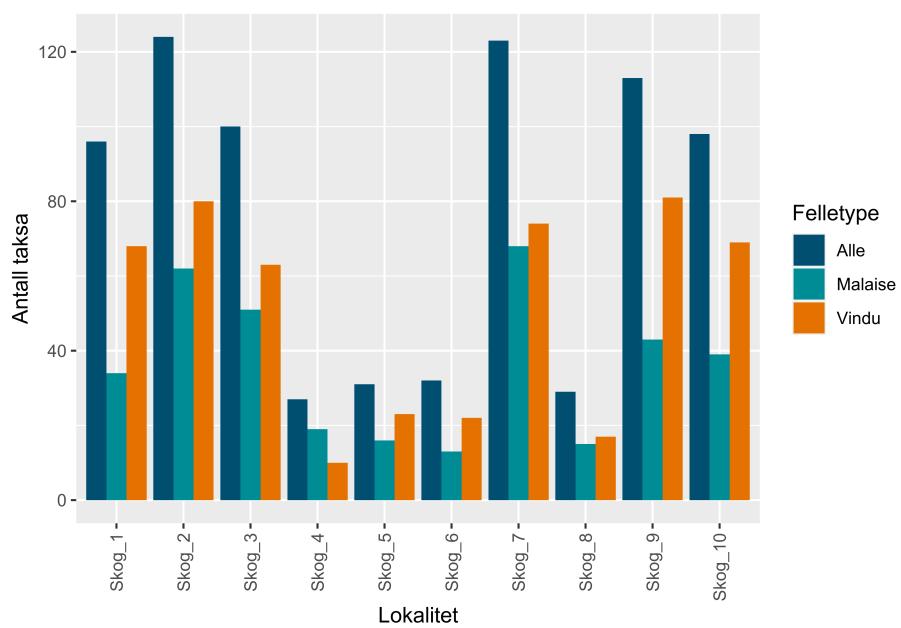


Figure 5: Cumulative number of Coleoptera taxa identified in each trap type

```

group_by(trap_type,
         genus,
         species_latin) %>%
summarize(no_catches = n())

## `summarise()` regrouping output by 'trap_type', 'genus' (override with `groups` argument)
p1 <- col_perc_traptyle %>%
  ungroup() %>%
  arrange(species_latin,
          trap_type) %>%
  filter(no_catches > 1) %>%
  slice(1:50) %>%
  group_by(trap_type) %>%
ggplot(.) +
  geom_bar(aes(y = reorder(species_latin, desc(species_latin)), x = no_catches, fill = trap_
                stat = "identity",
                position = position_dodge()) +
  scale_fill_nina(name = "Trap type") +
  ylab("") +
  xlab("Number of catches") +
  theme(legend.position = "none",
        text = element_text(size=5))

p2 <- col_perc_traptyle %>%
  ungroup() %>%
  arrange(species_latin,
          trap_type) %>%
  filter(no_catches > 1) %>%
  slice(51:100) %>%
  group_by(trap_type) %>%
ggplot(.) +
  geom_bar(aes(y = reorder(species_latin, desc(species_latin)), x = no_catches, fill = trap_
                stat = "identity",
                position = position_dodge()) +
  scale_fill_nina(name = "Trap type") +
  ylab("") +
  xlab("Number of catches") +
  theme(legend.position = "none",
        text = element_text(size=5))

p3 <- col_perc_traptyle %>%
  ungroup() %>%
  arrange(species_latin,
          trap_type) %>%

```

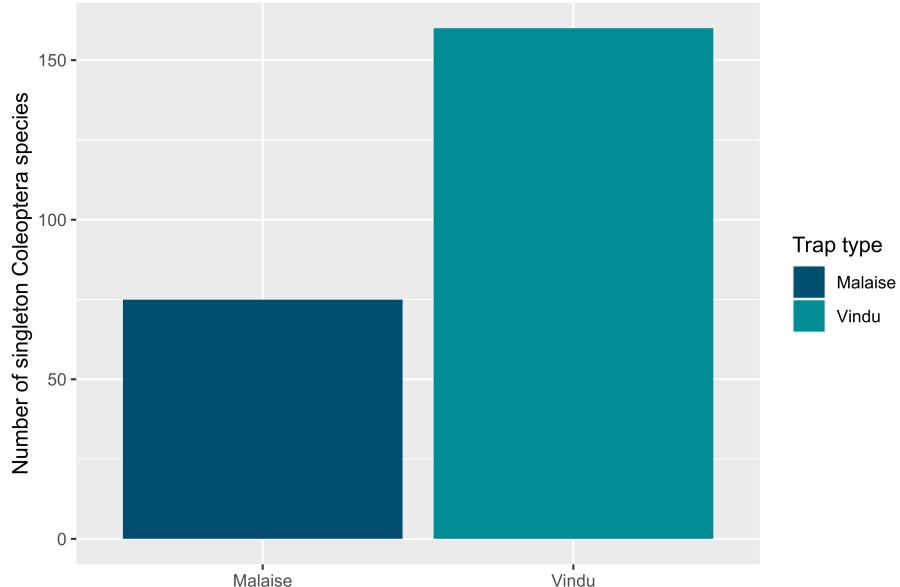


```

arrange(species_latin,
        trap_type) %>%
filter(no_catches == 1) %>%
group_by(trap_type) %>%
summarize(no_singletons = n()) %>%
ggplot(.) +
  geom_bar(aes(y = no_singletons, x = trap_type, fill = trap_type),
            stat = "identity") +
  scale_fill_nina(name = "Trap type") +
  ylab("Number of singleton Coleoptera species") +
  xlab("")

```

## `summarise()` ungrouping output (override with `.`groups` argument)



## Figs to Niclas on Pollinators

Niclas Gyllenstrand asked how efficient the malaisetraps+metabarcoding are at finding pollinators. This could be a relevant graph for the report as well.

```

poll_spec_richn <- dat %>%
  filter(trap_type == "Malaise") %>%
  filter(family == "Andrenidae" |
         family == "Apidae" |
         family == "Colletidae" |

```

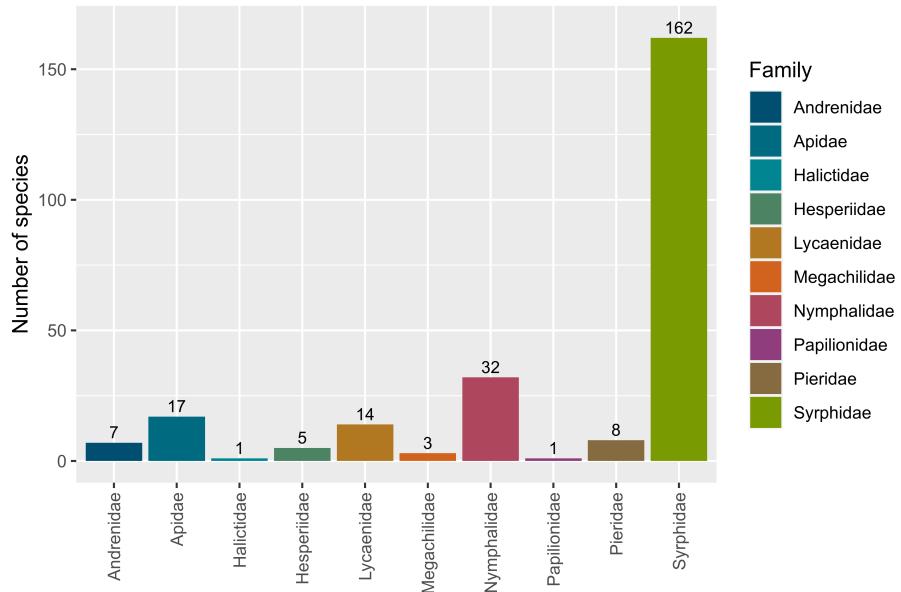
```

family == "Halictidae" |
family == "Megachilidae" |
family == "Melittidae" |
family == "Syrphidae" |

family == "Hesperiidae" |
family == "Lycaenidae" |
family == "Nymphalidae" |
family == "Papilionidae" |
family == "Pieridae" |
family == "Riodinidae"
) %>%
group_by(family) %>%
distinct(family, genus, species) %>%
summarize(col_spec_richn = n(),
no_unidentified_sp = sum(grepl("sp.", species)))

## `summarise()` ungrouping output (override with ` `.groups` argument)
ggplot(poll_spec_richn, aes(x = family, y = col_spec_richn, label = col_spec_richn)) +
  geom_bar(stat = "identity",
  aes(fill = family)) +
  geom_text(size = 3,
  position = position_nudge(y = 4)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  scale_fill_nina(name = "Family") +
  xlab("") +
  ylab("Number of species")

```



```

pollinator_species <- dat %>%
  filter(trap_type == "Malaise") %>%
  filter(family == "Andrenidae" |
         family == "Apidae" |
         family == "Colletidae" |
         family == "Halictidae" |
         family == "Megachilidae" |
         family == "Melittidae" |
         family == "Syrphidae" |

         family == "Hesperiidae" |
         family == "Lycaenidae" |
         family == "Nymphalidae" |
         family == "Papilionidae" |
         family == "Pieridae" |
         family == "Riodinidae"
     ) %>%
  distinct(order, family, genus, species) %>%
  arrange(family, genus, species)

dat %>%
  filter(trap_type == "Malaise") %>%
  filter(family == "Apidae") %>%
  distinct(genus, species) %>%
  select(genus, species)

```

```

## # A tibble: 17 x 2
##   genus     species
##   <chr>     <chr>
## 1 Apis      mellifera
## 2 Bombus    soroeensis
## 3 Bombus    pascuorum
## 4 Bombus    lucorum
## 5 Bombus    jonellus
## 6 Bombus    pratorum
## 7 Bombus    monticola
## 8 Melipona  illota
## 9 Bombus    consobrinus
## 10 Bombus   cryptarum
## 11 Bombus   sporadicus
## 12 Bombus   cingulatus
## 13 Bombus   hortorum
## 14 Bombus   hypnorum
## 15 Bombus   norvegicus
## 16 Bombus   ashtoni
## 17 Bombus   terrestris

write_csv(pollinator_species,
          path = "out/pollinator_species.csv")

```

## Phenology graphs

Some “timeseries” throughout the season.

```

seasonAvgNo <- dat %>%
  filter(weeks_sampled == 2) %>%
  group_by(habitat_type, date_emptied, locality) %>%
  summarise(no_species = n_distinct(genus, species),
            sum_biomass = sum(Vekt_flaske_.insekter_vaat_.g., na.rm = T) / no_species)

## `summarise()` regrouping output by 'habitat_type', 'date_emptied' (override with `.`groups)
# test <- seasonAvgNo <- dat %>%
#   filter(weeks_sampled == 2) %>%
#   group_by(habitat_type, date_emptied, locality) %>%
#   select(habitat_type, date_emptied, locality, Vekt_flaske_.insekter_vaat_.g.) %>%
#   distinct() %>%
#   summarise(sum(Vekt_flaske_.insekter_vaat_.g., na.rm = T))

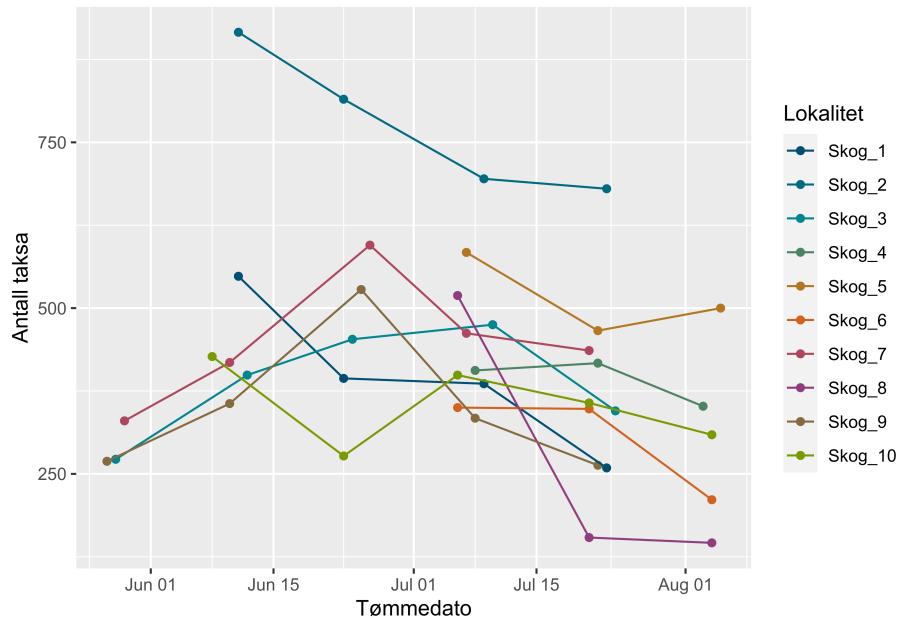
seasonAvgNo %>%
  filter(habitat_type == "Skog") %>%
  ggplot() +

```

```

geom_point(aes(x = dateemptied, y = no_species, color = locality)) +
geom_line(aes(x = dateemptied, y = no_species, color = locality)) +
ylab("Antall taksa") +
xlab("Tømmedato") +
scale_color_nina(name = "Lokalitet")

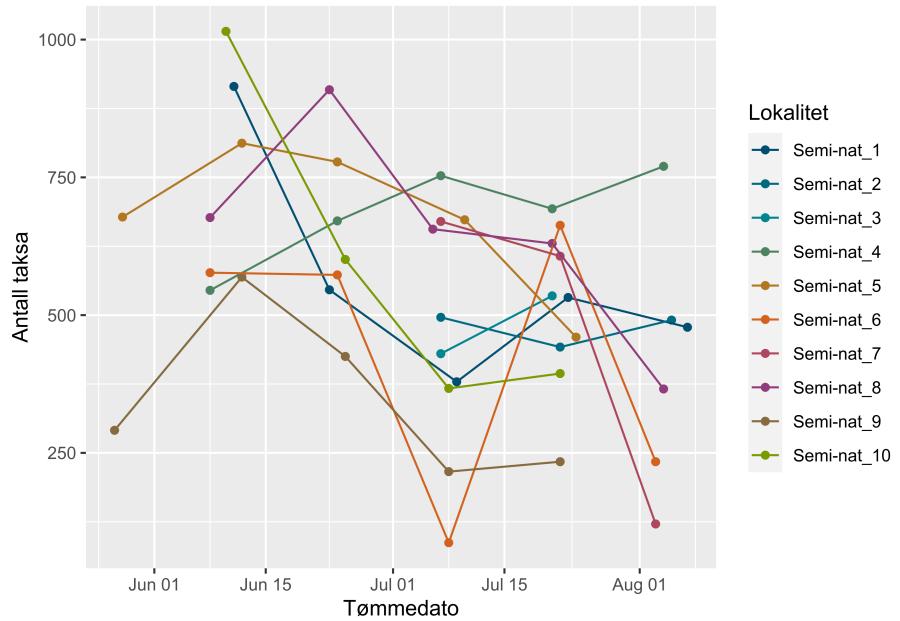
```



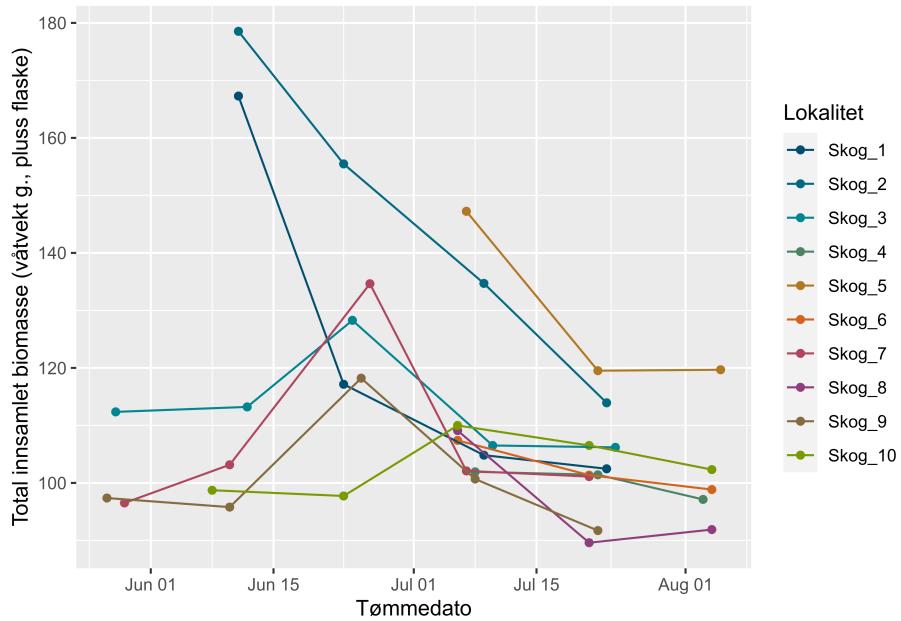
```

seasonAvgNo %>%
  filter(habitat_type == "SN_Gressmark") %>%
ggplot() +
  geom_point(aes(x = dateemptied, y = no_species, color = locality)) +
  geom_line(aes(x = dateemptied, y = no_species, color = locality)) +
  ylab("Antall taksa") +
  xlab("Tømmedato") +
  scale_color_nina(name = "Lokalitet")

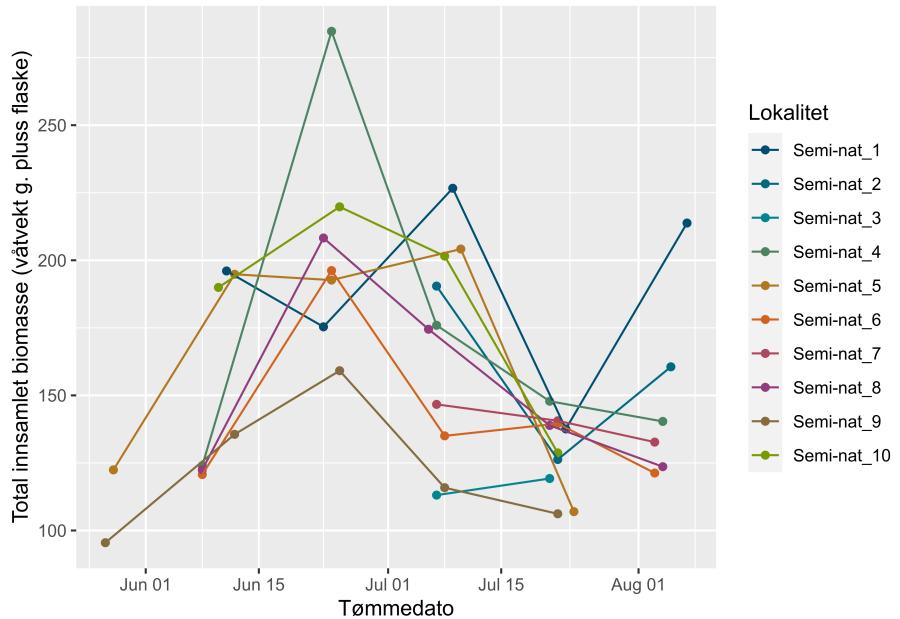
```



```
seasonAvgNo %>%
  filter(habitat_type == "Skog") %>%
  ggplot() +
  geom_point(aes(x = dateemptied, y = sum_biomass, color = locality)) +
  geom_line(aes(x = dateemptied, y = sum_biomass, color = locality)) +
  ylab("Total innsamlet biomasse (våtvekt g., pluss flaske)") +
  xlab("Tømmedato") +
  scale_color_nina(name = "Lokalitet")
```



```
seasonAvgNo %>%
  filter(habitat_type == "SN_Gressmark") %>%
  ggplot() +
  geom_point(aes(x = dateemptied, y = sum_biomass, color = locality)) +
  geom_line(aes(x = dateemptied, y = sum_biomass, color = locality)) +
  ylab("Total innsamlet biomasse (våtvekt g. pluss flaske)") +
  xlab("Tømmedato") +
  scale_color_nina(name = "Lokalitet")
```



## Plot some weather data

```
logger_tbl <- tbl(con,
                   in_schema("loggers", "logger_data")) %>%
  select(-id)

logger_dep_tbl <- tbl(con,
                      in_schema("loggers", "logger_deployments")) %>%
  select(-id)

logger_data <- logger_tbl %>%
  left_join(logger_dep_tbl,
            by = c("logger_id" = "logger_id",
                  "logger_type" = "logger_type")) %>%
  mutate(day = as.Date(date))
```

Plot average values for each site during July

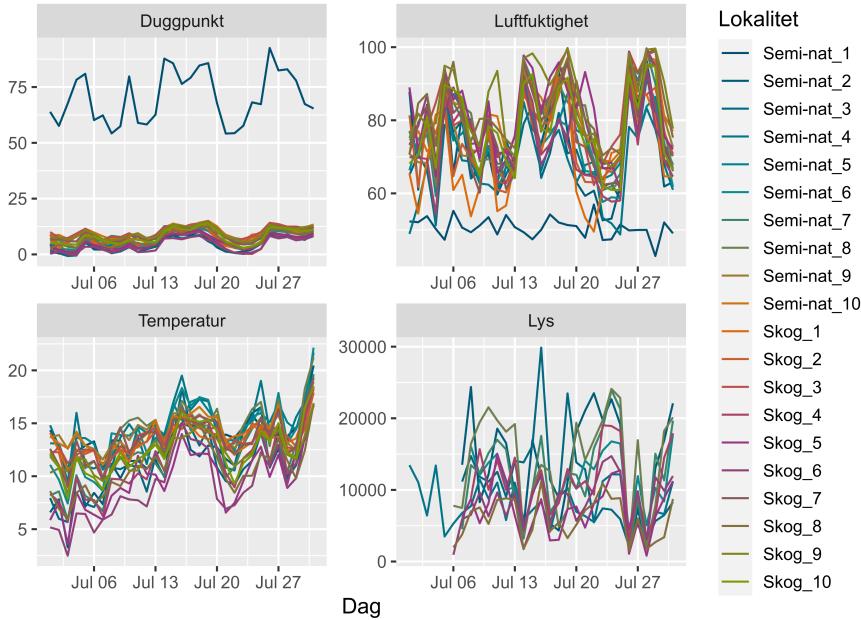
```
tt <- logger_data %>%
  group_by(location, day, data_type) %>%
  summarize(value = mean(value)) %>%
  filter(day >= '2020-07-01' &
         day < '2020-08-01'
```

```

) %>%
collect() %>%
mutate(Lokalitet = factor(location, levels = c(paste0("Semi-nat_", 1:10), paste0("Skog_", 1:10)),
  mutate(data_type = factor(data_type, levels = )))
##Why can't I set the levels in the pipe?
levels(tt$data_type) <- c("Duggpunkt", "Luftfuktighet", "Temperatur", "Lys")

tt %>%
  ggplot(.) +
  geom_line(aes(x = day, y = value, color = Lokalitet)) +
  facet_wrap(~data_type,
             scales = "free") +
  scale_color_nina(name = "Lokalitet") +
  xlab("Dag") +
  ylab("") +
  theme(legend.key.height = unit(0.5, "cm"))

```



There seems to be two clusters of relative humidity values. In this updated data, we see that only the Semi-nat\_1 logger is off. That is the only one where we didn't download the data through Hobolink. So it seems that something happens when we run the data through hobolink, compared to when we download the data in a csv file directly from the logger. Must talk to Hobo about this.

```

location_groups <- tibble(location = c(paste0("Semi-nat_", 1:10), paste0("Skog_", 1:10)),
  location_group = c("Oslo", "Lillehammer", "Lillehammer", "Lilleham")

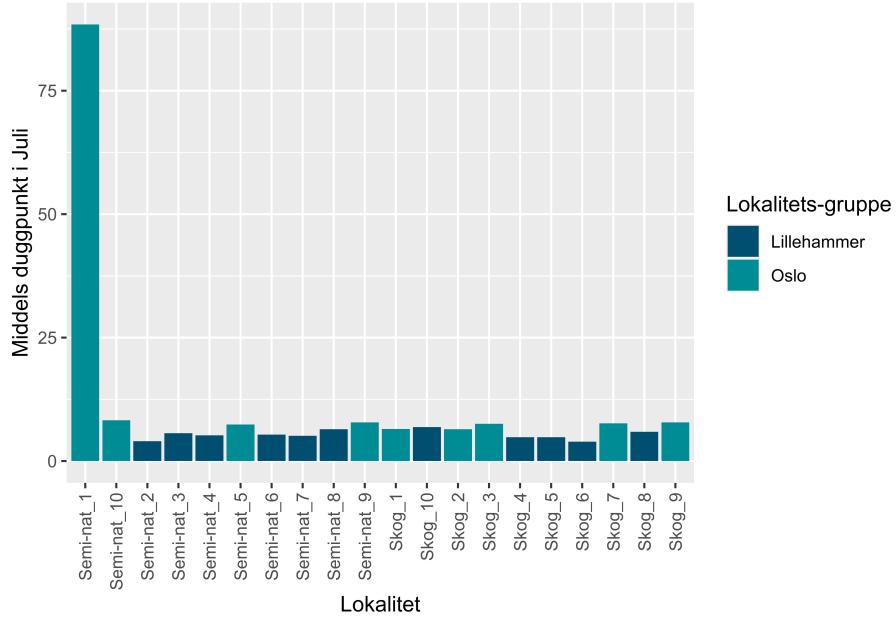
```

```

    "Oslo", "Lillehammer", "Lillehammer", "Lilleham
    "Oslo", "Oslo", "Oslo", "Oslo", "Oslo", "Lille
    "Lillehammer", "Lillehammer", "Oslo", "Lilleham
    "Oslo", "Lillehammer"))

logger_data %>%
  filter(day >= '2020-10-01' &
        day < '2020-11-01') %>%
  group_by(location, data_type) %>%
  summarize(value = mean(value)) %>%
  filter(data_type == "dew") %>%
  collect() %>%
  mutate(location = factor(location, levels = c(paste0("Semi-nat_", 1:10), paste0("Skog_", 1:9)),
    left_join(location_groups,
      by = c("location" = "location")) %>%
  ggplot(.) +
  geom_bar(aes(x = location, y = value, fill = location_group), stat = "identity") +
  scale_fill_nina(name = "Lokalitets-gruppe") +
  xlab("Lokalitet") +
  ylab("Middels duggpunkt i Juli") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)))

```



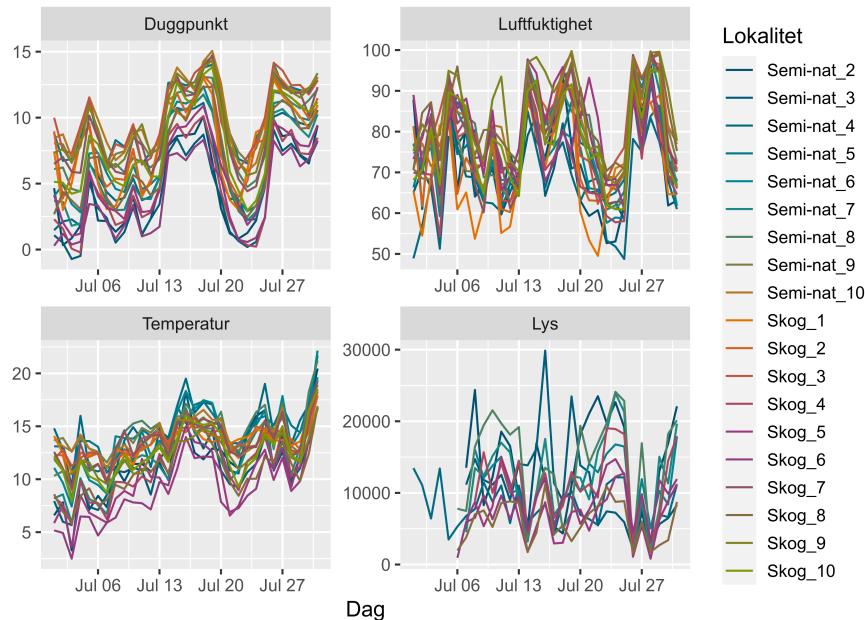
Plot all but the Semi-nat\_1 data.

```

tt <- logger_data %>%
  filter(location != "Semi-nat_1") %>%
  group_by(location, day, data_type) %>%
  summarize(value = mean(value)) %>%
  filter(day >= '2020-07-01' &
        day < '2020-08-01'
      ) %>%
  collect() %>%
  mutate(Lokalitet = factor(location, levels = c(paste0("Semi-nat_", 1:10), paste0("Skog_", 1:10)),
                            data_type = factor(data_type, levels = )))
##Why can't I set the levels in the pipe?
levels(tt$data_type) <- c("Duggpunkt", "Luftfuktighet", "Temperatur", "Lys")

tt %>%
  ggplot(.) +
  geom_line(aes(x = day, y = value, color = Lokalitet)) +
  facet_wrap(~data_type,
             scales = "free") +
  scale_color_nina(name = "Lokalitet") +
  xlab("Dag") +
  ylab("") +
  theme(legend.key.height = unit(0.5, "cm"))

```



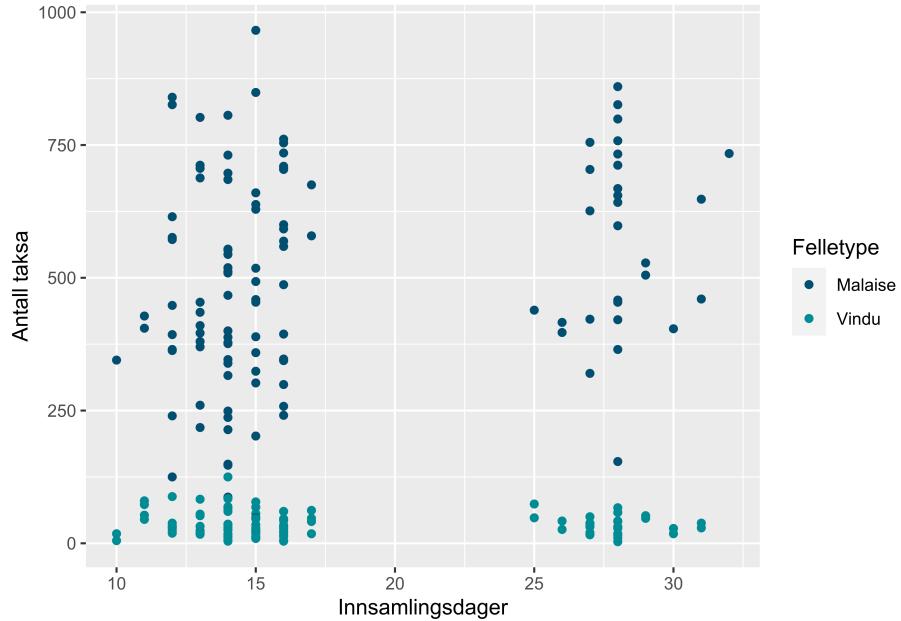
## Sampling time analysis

Time to look at the impact of sampling time, 2 vs 4 weeks.

First off, simple plot of species richness against days sampled.

```
dat %>%
  group_by(locality,
    trap,
    trap_type,
    empty_week,
    weeks_sampled,
    days_collected) %>%
  summarise(no_spec = n_distinct(order, family, genus, species)) %>%
  ggplot(aes(x = days_collected, y = no_spec)) +
  geom_point(aes(color = trap_type)) +
  scale_color_nina(name = "Felletype") +
  ylab("Antall taksa") +
  xlab("Innsamlingsdager")
```

## `summarise()` regrouping output by 'locality', 'trap', 'trap\_type', 'empty\_week', 'weeks\_



Not easy to see any patterns here, since there is such variation in the catches throughout the season. We need to sum up the catches in the 2 2-week periods and compare to the single 4 week periods.

We first identify which of the 2-week samples that should be aggregated.

```

#4-week "target" to agg to
week_4 <- dat %>%
  filter(weeks_sampled == 4) %>%
  select(locality,
         empty_week) %>%
  distinct() %>%
  mutate(agg_id = 1:n())

#add 2 weeks earlier
earlier_2_weeks <- week_4 %>%
  select(locality,
         agg_id) %>%
  mutate(empty_week = week_4$empty_week - 2)

which_to_agg <- week_4 %>%
  bind_rows(earlier_2_weeks)

#add aggregation ids to 2-week samplings
to_agg <- dat %>%
  filter(weeks_sampled == 2) %>%
  left_join(which_to_agg,
            by = c("locality" = "locality",
                  "empty_week" = "empty_week"))

```

Check which didn't get an aggregation

```

to_agg %>%
  filter(is.na(agg_id)) %>%
  select(locality,
         empty_week,
         trap) %>%
  distinct() %>%
  arrange(empty_week,
         locality) %>%
  print(n = Inf)

## # A tibble: 46 x 3
##   locality    empty_week trap
##   <fct>        <dbl> <chr>
## 1 Semi-nat_1     24  MF1
## 2 Semi-nat_1     24  MF2
## 3 Semi-nat_10    24  MF1
## 4 Semi-nat_10    24  MF2
## 5 Skog_1          24  MF1
## 6 Skog_1          24  MF2
## 7 Skog_1          24  VF1

```

```

##  8 Skog_1          24 VF2
##  9 Skog_1          24 VF3
## 10 Skog_1          24 VF4
## 11 Skog_2          24 MF1
## 12 Skog_2          24 MF2
## 13 Skog_2          24 VF1
## 14 Skog_2          24 VF2
## 15 Skog_2          24 VF3
## 16 Skog_2          24 VF4
## 17 Semi-nat_1      30 MF1
## 18 Semi-nat_5      30 MF1
## 19 Semi-nat_9      30 MF1
## 20 Semi-nat_10     30 MF1
## 21 Skog_1          30 MF1
## 22 Skog_1          30 VF1
## 23 Skog_1          30 VF3
## 24 Skog_2          30 MF1
## 25 Skog_2          30 VF1
## 26 Skog_2          30 VF3
## 27 Skog_3          30 MF1
## 28 Skog_3          30 VF1
## 29 Skog_3          30 VF3
## 30 Skog_7          30 MF1
## 31 Skog_7          30 VF1
## 32 Skog_7          30 VF3
## 33 Skog_9          30 MF1
## 34 Skog_9          30 VF1
## 35 Skog_9          30 VF3
## 36 Semi-nat_1      32 MF1
## 37 Semi-nat_2      32 MF1
## 38 Semi-nat_4      32 MF1
## 39 Semi-nat_6      32 MF1
## 40 Semi-nat_7      32 MF1
## 41 Semi-nat_8      32 MF1
## 42 Skog_4          32 MF1
## 43 Skog_5          32 MF1
## 44 Skog_6          32 MF1
## 45 Skog_8          32 MF1
## 46 Skog_10         32 MF1

```

The samples from week 24 are expected, since not all traps where set up early enough to get 2 emptyings by then.

```

dat %>%
  filter(locality == 'Semi-nat_2' &
        empty_week == 34) %>%

```

```

    select(locality,
           empty_week,
           dateemptied,
           weeks_sampled,
           trap,
           GenlabID) %>%
  distinct()

## # A tibble: 0 x 6
## # ... with 6 variables: locality <fct>, empty_week <dbl>, dateemptied <date>,
## #   weeks_sampled <dbl>, trap <chr>, GenlabID <chr>

But Semi-nat_1, week_32, trap MF2 should be there. Semi-nat_5 as well.
Seems that these haven't been run yet. I'll remove these for the comparison.

#Sum the genetic reads to combine the two weeks
agg_data_2_week <- to_agg %>%
  filter(!is.na(agg_id)) %>%
  group_by(locality,
           trap_type,
           trap,
           agg_id,
           species_latin) %>%
  summarise(value = sum(value)) %>%
  ungroup()

## `summarise()` regrouping output by 'locality', 'trap_type', 'trap', 'agg_id' (override w
#Add the week information for the second sampling
agg_data_2_week <- agg_data_2_week %>%
  left_join(week_4,
            by = c("agg_id" = "agg_id",
                  "locality" = "locality")) %>%
  mutate(weeks_sampled = 2) %>%
  select(locality,
         empty_week,
         weeks_sampled,
         trap_type,
         trap,
         species_latin,
         value)

#Select only the original 4 week samplings
data_4_week <- dat %>%
  filter(weeks_sampled == 4) %>%
  select(locality,

```

```

    empty_week,
    weeks_sampled,
    trap_type,
    trap,
    species_latin,
    value)

#Add the original 4 weeks to the summarized 2 2-week samplings
week_comp_data <- agg_data_2_week %>%
  bind_rows(data_4_week)

```

Summarise the number of species in both sample lengths.

```

no_species_weeks <- week_comp_data %>%
  group_by(locality,
           empty_week,
           weeks_sampled,
           trap_type) %>%
  summarise(no_species = n())

## `summarise()` regrouping output by 'locality', 'empty_week', 'weeks_sampled' (override with `.`groups` argument)
#Check comparison
no_species_weeks %>%
  group_by(locality,
           empty_week,
           trap_type) %>%
  summarise(no_comp = n()) %>%
  print(n = Inf)

## `summarise()` regrouping output by 'locality', 'empty_week' (override with `.`groups` argument)
## # A tibble: 43 x 4
## # Groups:   locality, empty_week [29]
##   locality   empty_week trap_type no_comp
##   <fct>       <dbl> <chr>      <int>
## 1 Semi-nat_1     28 Malaise      2
## 2 Semi-nat_2     30 Malaise      2
## 3 Semi-nat_3     30 Malaise      2
## 4 Semi-nat_4     26 Malaise      2
## 5 Semi-nat_4     30 Malaise      2
## 6 Semi-nat_5     24 Malaise      2
## 7 Semi-nat_5     28 Malaise      2
## 8 Semi-nat_6     26 Malaise      2
## 9 Semi-nat_6     30 Malaise      2
## 10 Semi-nat_7    30 Malaise      2
## 11 Semi-nat_8    26 Malaise      2

```

```

## 12 Semi-nat_8      30 Malaise    2
## 13 Semi-nat_9      24 Malaise    2
## 14 Semi-nat_9      28 Malaise    2
## 15 Semi-nat_10     28 Malaise    2
## 16 Skog_1           28 Malaise    2
## 17 Skog_1           28 Vindu      2
## 18 Skog_2           28 Malaise    2
## 19 Skog_2           28 Vindu      2
## 20 Skog_3           24 Malaise    2
## 21 Skog_3           24 Vindu      2
## 22 Skog_3           28 Malaise    2
## 23 Skog_3           28 Vindu      2
## 24 Skog_4           30 Malaise    2
## 25 Skog_4           30 Vindu      2
## 26 Skog_5           30 Malaise    2
## 27 Skog_5           30 Vindu      2
## 28 Skog_6           30 Malaise    2
## 29 Skog_6           30 Vindu      2
## 30 Skog_7           24 Malaise    2
## 31 Skog_7           24 Vindu      2
## 32 Skog_7           28 Malaise    2
## 33 Skog_7           28 Vindu      2
## 34 Skog_8           30 Malaise    2
## 35 Skog_8           30 Vindu      2
## 36 Skog_9           24 Malaise    2
## 37 Skog_9           24 Vindu      2
## 38 Skog_9           28 Malaise    2
## 39 Skog_9           28 Vindu      2
## 40 Skog_10          26 Malaise    2
## 41 Skog_10          26 Vindu      2
## 42 Skog_10          30 Malaise    2
## 43 Skog_10          30 Vindu      2

```

Make a comparison table.

```

comp_species_week <- no_species_weeks %>%
  pivot_wider(id_cols = c(locality,
                           empty_week,
                           trap_type),
              names_from = weeks_sampled,
              values_from = no_species,
              names_prefix = "sample_time_") %>%
  mutate(diff_sample_time = sample_time_4 - sample_time_2,
        prc_diff_sample_time = (sample_time_4 - sample_time_2) / sample_time_2)

```

Quality check of a few rows. Seems OK.

```

dat %>%
  filter(locality == "Semi-nat_1" &
         (empty_week == 26 | empty_week == 28) &
         trap == "MF1"
    ) %>%
  summarize(no_spec = n_distinct(species_latin))

## # A tibble: 1 x 1
##   no_spec
##       <int>
## 1      776

dat %>%
  filter(locality == "Semi-nat_5" &
         (empty_week == 22 | empty_week == 24) &
         trap == "MF1"
    ) %>%
  summarize(no_spec = n_distinct(species_latin))

## # A tibble: 1 x 1
##   no_spec
##       <int>
## 1      1219

dat %>%
  filter(locality == "Semi-nat_5" &
         (empty_week == 24 ) &
         trap == "MF2"
    ) %>%
  summarize(no_spec = n_distinct(species_latin))

## # A tibble: 1 x 1
##   no_spec
##       <int>
## 1      648

dat %>%
  filter(locality == "Semi-nat_6" &
         (empty_week == 26 ) &
         trap == "MF2"
    ) %>%
  summarize(no_spec = n_distinct(species_latin))

## # A tibble: 1 x 1
##   no_spec
##       <int>
## 1      826

```

```

week_comp_data %>%
  filter(locality == "Semi-nat_5" &
         (empty_week == 24) &
         trap == "MF2"
    ) %>%
  summarize(no_spec = n_distinct(species_latin))

## # A tibble: 1 x 1
##   no_spec
##   <int>
## 1     648

no_species_weeks %>%
  filter(locality == "Semi-nat_5" &
         empty_week == 24)

## # A tibble: 2 x 5
## # Groups:   locality, empty_week, weeks_sampled [2]
##   locality   empty_week weeks_sampled trap_type no_species
##   <fct>        <dbl>      <dbl> <chr>       <int>
## 1 Semi-nat_5      24          2 Malaise      1219
## 2 Semi-nat_5      24          4 Malaise       648

comp_species_week %>%
  group_by(locality) %>%
  filter(trap_type == "Malaise") %>%
  ggplot(.) +
  geom_point(aes(x = empty_week, y = diff_sample_time, color = locality)) +
  geom_line(aes(x = empty_week, y = diff_sample_time, color = locality)) +
  scale_color_nina(name = "Lokalitet") +
  ylab("Diff. kumulativ 2 uker - 4 uker (no. taksa)") +
  xlab("Tømmingsuke, 4 ukerstømming") +
  geom_hline(yintercept = 0) +
  theme(legend.key.height = unit(0.5, "cm"))

comp_species_week %>%
  group_by(locality) %>%
  filter(trap_type == "Malaise") %>%
  ggplot(.) +
  geom_point(aes(x = empty_week, y = prc_diff_sample_time, color = locality)) +
  geom_line(aes(x = empty_week, y = prc_diff_sample_time, color = locality)) +
  scale_color_nina(name = "Lokalitet") +
  ylab("% Diff. kumulativ 2 uker - 4 uker") +
  xlab("Tømmingsuke, 4 ukerstømming") +
  geom_hline(yintercept = 0) +
  theme(legend.key.height = unit(0.5, "cm"))

```

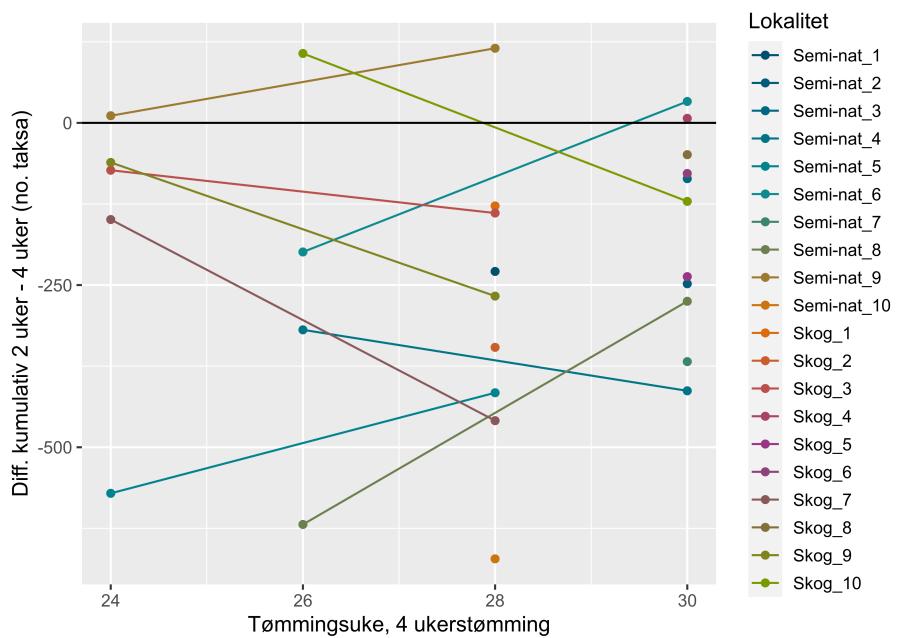


Figure 6: Difference (g.) in cumulative species richness over 2 2-week samplings, compared to 1 4 week sampling. Catch in 1 Malaise trap per period. Negative values indicate that 4-week samplings gave lower insect species richness than the two 2-week samplings.

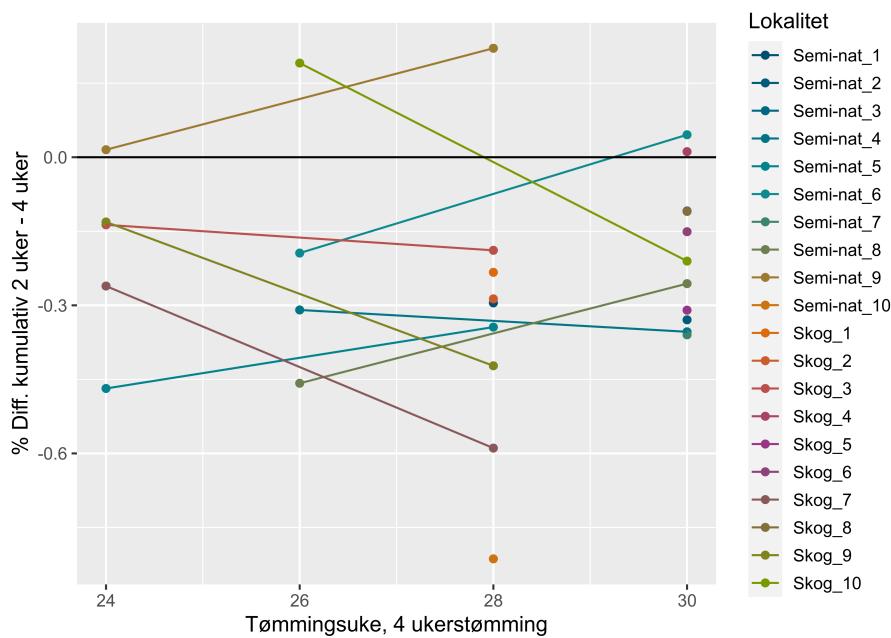


Figure 7: Percentage difference in cumulative species richness over 2 2-week samplings, compared to 1 4 week sampling. Catch in 1 Malaise trap per period. Negative values indicate that 4-week samplings gave lower insect species richness than the two 2-week samplings.

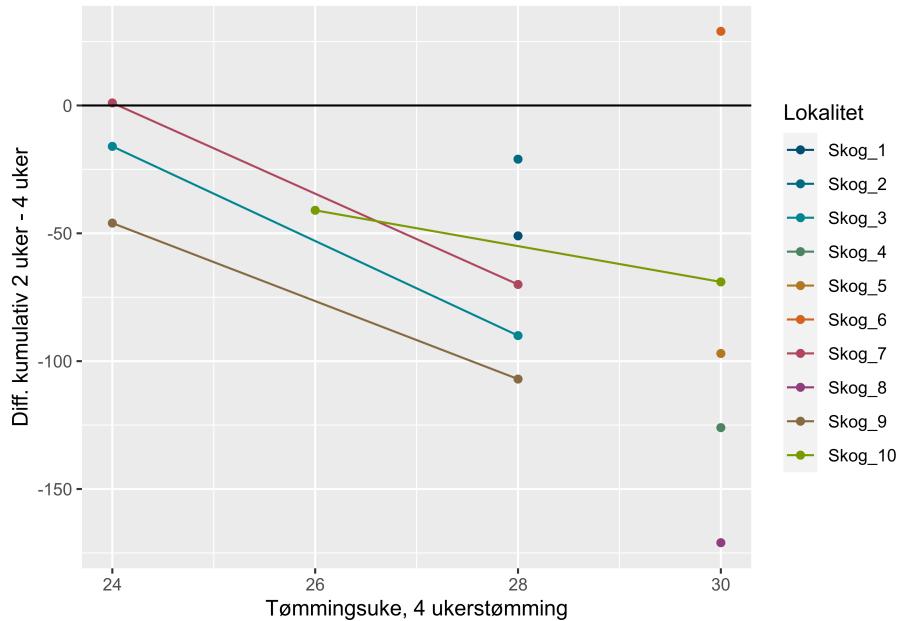


Figure 8: Difference in cumulative species richness over 2 2-week samplings, compared to 1 4 week sampling. Catch in 1 Malaise trap per period. Negative values indicate that 4-week samplings gave lower insect species richness than the two 2-week samplings.

```

comp_species_week %>%
  group_by(locality) %>%
  filter(trap_type == "Vindu") %>%
  ggplot(.) +
  geom_point(aes(x = empty_week, y = diff_sample_time, color = locality)) +
  geom_line(aes(x = empty_week, y = diff_sample_time, color = locality)) +
  scale_color_nina(name = "Lokalitet") +
  ylab("Diff. kumulativ 2 uker - 4 uker") +
  xlab("Tømmingsuke, 4 ukerstømning") +
  geom_hline(yintercept = 0)

comp_species_week %>%
  group_by(locality) %>%
  filter(trap_type == "Vindu") %>%
  ggplot(.) +
  geom_point(aes(x = empty_week, y = prc_diff_sample_time, color = locality)) +
  geom_line(aes(x = empty_week, y = prc_diff_sample_time, color = locality)) +
  scale_color_nina(name = "Lokalitet") +
  ylab("% Diff. kumulativ 2 uker - 4 uker") +

```

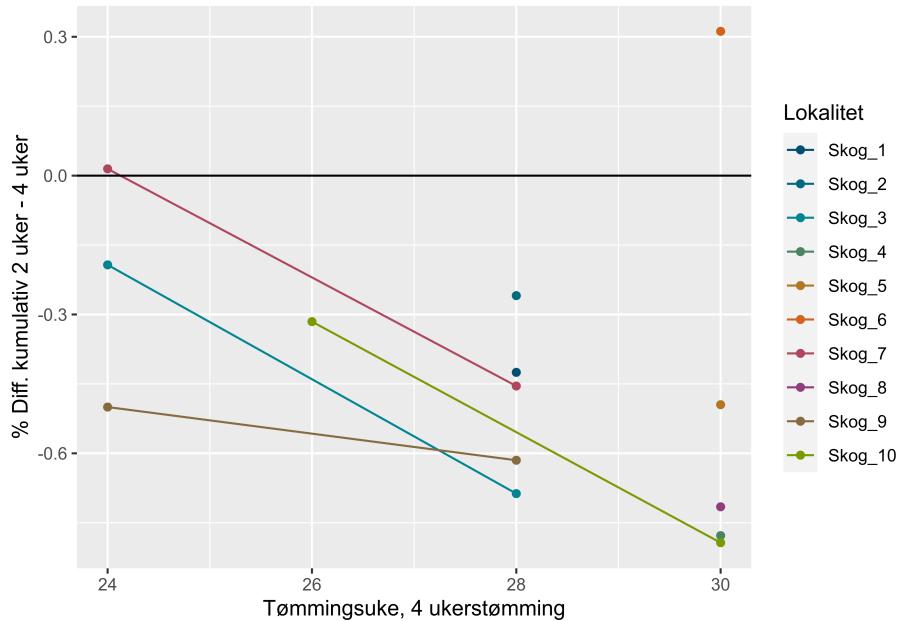


Figure 9: Percentage difference in cumulative species richness over 2 2-week samplings, compared to 1 4 week sampling. Catch in 2 window traps per period. Negative values indicate that 4-week samplings gave lower insect species richness than the two 2-week samplings.

```
xlab("Tømmingsuke, 4 ukerstømming") +
  geom_hline(yintercept = 0)
```

### A similar comparison, but for biomass.

```
biomass_agg_data_2_week <- to_agg %>%
  filter(!is.na(agg_id)) %>%
  select(locality,
         empty_week,
         trap_type,
         trap,
         agg_id,
         biomass = Toerrvekt_.g.) %>%
  distinct() %>%
  arrange(locality,
         empty_week,
         trap) %>%
```

```

group_by(locality,
         trap_type,
         agg_id) %>%
summarise(empty_week = max(empty_week),
           biomass = sum(biomass)
         ) %>%
ungroup() %>%
select(-agg_id)

## `summarise()` regrouping output by 'locality', 'trap_type' (override with `.`groups` argument)
biomass_data_4_week <- dat %>%
  filter(weeks_sampled == 4) %>%
  select(locality,
         empty_week,
         trap_type,
         trap,
         biomass = Toerrvekt_.g.) %>%
distinct() %>%
arrange(locality,
        empty_week,
        trap) %>%
group_by(locality,
         empty_week,
         trap_type) %>%
summarise(empty_week = max(empty_week),
           biomass = sum(biomass)
         ) %>%
ungroup()

## `summarise()` regrouping output by 'locality', 'empty_week' (override with `.`groups` argument)
#Make comparison table
biomass_comp <- biomass_agg_data_2_week %>%
  left_join(biomass_data_4_week,
            by = c("locality" = "locality",
                  "empty_week" = "empty_week",
                  "trap_type" = "trap_type"),
            suffix = c("_2_weeks", "_4_weeks")) %>%
  mutate(diff_sample_time = biomass_4_weeks - biomass_2_weeks,
         prc_diff_sample_time = (biomass_4_weeks - biomass_2_weeks) / biomass_2_weeks)

biomass_comp %>%
  group_by(locality) %>%
  filter(trap_type == "Malaise") %>%
  ggplot(.) +
  geom_point(aes(x = empty_week, y = prc_diff_sample_time, color = locality)) +

```

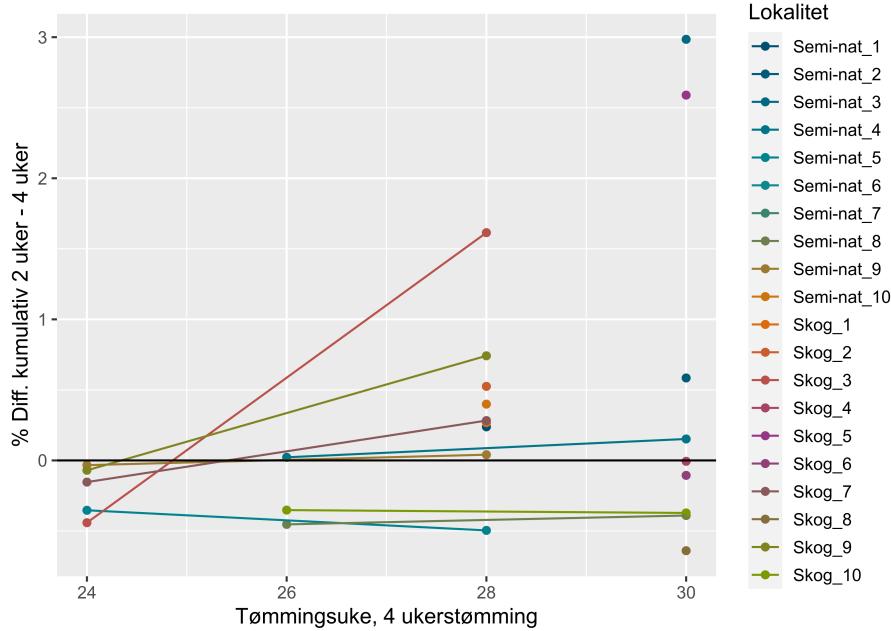


Figure 10: Percentage difference in cumulative dry weight over 2 2-week samplings, compared to 1 4 week sampling. Catch in 1 Malaise trap per period. Negative values indicate that 4-week samplings gave lower insect species richness than the two 2-week samplings.

```
geom_line(aes(x = empty_week, y = prc_diff_sample_time, color = locality)) +
  scale_color_nina(name = "Lokalitet") +
  ylab("% Diff. kumulativ 2 uker - 4 uker") +
  xlab("Tømmingsuke, 4 ukerstømning") +
  geom_hline(yintercept = 0) +
  theme(legend.key.height = unit(0.5, "cm"))
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
## Warning: Removed 3 row(s) containing missing values (geom_path).
```

We can do the same comparison with wet weight, although we might expect that this will add some weight of the ethanol and the bottles, which will lead to bias when we sum these together for the 2 weeks. Anyway, this is the way it looks for wet weight.

```
biomass_agg_data_2_week <- to_agg %>%
  filter(!is.na(agg_id)) %>%
  select(locality,
         empty_week,
```

```

    trap_type,
    trap,
    agg_id,
    biomass = Vaatvekt_.g.) %>%
distinct() %>%
arrange(locality,
        empty_week,
        trap) %>%
group_by(locality,
          trap_type,
          agg_id) %>%
summarise(empty_week = max(empty_week),
           biomass = sum(biomass)
           ) %>%
ungroup() %>%
select(-agg_id)

## `summarise()` regrouping output by 'locality', 'trap_type' (override with `.`groups` argument)
biomass_data_4_week <- dat %>%
  filter(weeks_sampled == 4) %>%
  select(locality,
         empty_week,
         trap_type,
         trap,
         biomass = Vaatvekt_.g.) %>%
distinct() %>%
arrange(locality,
        empty_week,
        trap) %>%
group_by(locality,
          empty_week,
          trap_type) %>%
summarise(empty_week = max(empty_week),
           biomass = sum(biomass)
           ) %>%
ungroup()

## `summarise()` regrouping output by 'locality', 'empty_week' (override with `.`groups` argument)
#Make comparison table
biomass_comp <- biomass_agg_data_2_week %>%
  left_join(biomass_data_4_week,
            by = c("locality" = "locality",
                  "empty_week" = "empty_week",
                  "trap_type" = "trap_type"),
            suffix = c("_2_weeks", "_4_weeks")) %>%

```

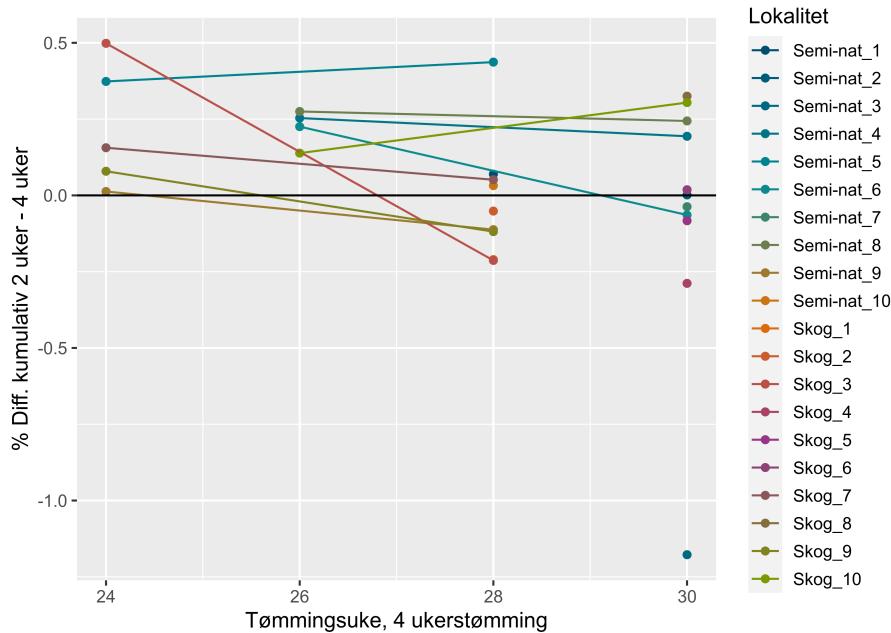


Figure 11: Percentage difference in cumulative wet weight over 2 2-week samplings, compared to 1 4 week sampling. Catch in 1 Malaise trap per period. Negative values indicate that 4-week samplings gave lower insect species richness than the two 2-week samplings.

```

  mutate(diff_sample_time = biomass_2_weeks - biomass_4_weeks,
         prc_diff_sample_time = (biomass_2_weeks - biomass_4_weeks) / biomass_2_weeks)

biomass_comp %>%
  group_by(locality) %>%
  filter(trap_type == "Malaise") %>%
  ggplot(.) +
  geom_point(aes(x = empty_week, y = prc_diff_sample_time, color = locality)) +
  geom_line(aes(x = empty_week, y = prc_diff_sample_time, color = locality)) +
  scale_color_nina(name = "Lokalitet") +
  ylab("% Diff. kumulativ 2 uker - 4 uker") +
  xlab("Tømmingsuke, 4 ukerstømming") +
  geom_hline(yintercept = 0) +
  theme(legend.key.height = unit(0.5, "cm"))

```

## Compare the ethanol to the propglykol window traps

```
liquid_comp <- dat %>%
  filter(trap_type == "Vindu") %>%
  group_by(locality,
    trap,
    liquid,
    weeks_sampled,
    empty_week,
    sample_id
  ) %>%
  summarise(no_species = n_distinct(species_latin))

## `summarise()` regrouping output by 'locality', 'trap', 'liquid', 'weeks_sampled', 'empty_...
liquid_comp

## # A tibble: 101 x 7
## # Groups:   locality, trap, liquid, weeks_sampled, empty_week [101]
##   locality trap   liquid   weeks_sampled empty_week sample_id no_species
##   <fct>    <chr> <chr>      <dbl>        <dbl> <chr>       <int>
## 1 Skog_1   VF1   Ethanol     2           24 skog_1_VF1_we~     43
## 2 Skog_1   VF1   Ethanol     2           26 skog_1_VF1_we~     31
## 3 Skog_1   VF1   Ethanol     2           28 skog_1_VF1_we~     30
## 4 Skog_1   VF1   Ethanol     2           30 skog_1_VF1_we~     27
## 5 Skog_1   VF2   Ethanol     2           24 skog_1_VF2_we~     60
## 6 Skog_1   VF2   Ethanol     4           28 skog_1_VF2_we~     28
## 7 Skog_1   VF3   PropGl_Eth~     2           24 skog_1_VF3_we~     46
## 8 Skog_1   VF3   PropGl_Eth~     2           26 skog_1_VF3_we~     31
## 9 Skog_1   VF3   PropGl_Eth~     2           28 skog_1_VF3_we~     46
## 10 Skog_1  VF3   PropGl_Eth~     2           30 skog_1_VF3_we~     18
## # ... with 91 more rows

liquid_mod_0 <- glmer(no_species ~ liquid * weeks_sampled + (1|empty_week) + (1 | locality),
                       family = "poisson",
                       data = liquid_comp)

isSingular(liquid_mod_0)

## [1] FALSE

summary(liquid_mod_0)

## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: poisson  ( log )
```

```

## Formula: no_species ~ liquid * weeks_sampled + (1 | empty_week) + (1 |
##   locality)
## Data: liquid_comp
##
##      AIC      BIC logLik deviance df.resid
## 1620.7 1636.4 -804.4  1608.7      95
##
## Scaled residuals:
##   Min     1Q Median     3Q    Max
## -5.8003 -2.5318 -0.8503  2.0783 10.6207
##
## Random effects:
## Groups   Name        Variance Std.Dev.
## locality (Intercept) 0.08466  0.2910
## empty_week (Intercept) 0.05510  0.2347
## Number of obs: 101, groups: locality, 10; empty_week, 5
##
## Fixed effects:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                3.62137   0.15886 22.797 <2e-16 ***
## liquidPropGl_Ethanol      -0.03270   0.10194 -0.321  0.748
## weeks_sampled              -0.04275   0.02861 -1.494  0.135
## liquidPropGl_Ethanol:weeks_sampled 0.04058   0.03826  1.061  0.289
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) lqPG_E wks_sm
## lqdPrpGl_Et -0.342
## weeks_smpld -0.438  0.679
## lqdPrpG_E:_  0.328 -0.944 -0.728

```

I struggle to model this with all the random effect I would like. I'm not able to include a id level random effect to account for overdispersion for example. Anyway, the results give an indication of the differences in the data.

```

liquid_pred <- ggpredict(liquid_mod_0,
                           terms = c("liquid", "weeks_sampled"))

plot(liquid_pred) +
  scale_color_nina(name = "Antall uker") +
  ggtitle("") +
  ylab("Antall taksa") +
  xlab("Væske i vindusfelle")

## Scale for 'colour' is already present. Adding another scale for 'colour',
## which will replace the existing scale.

```

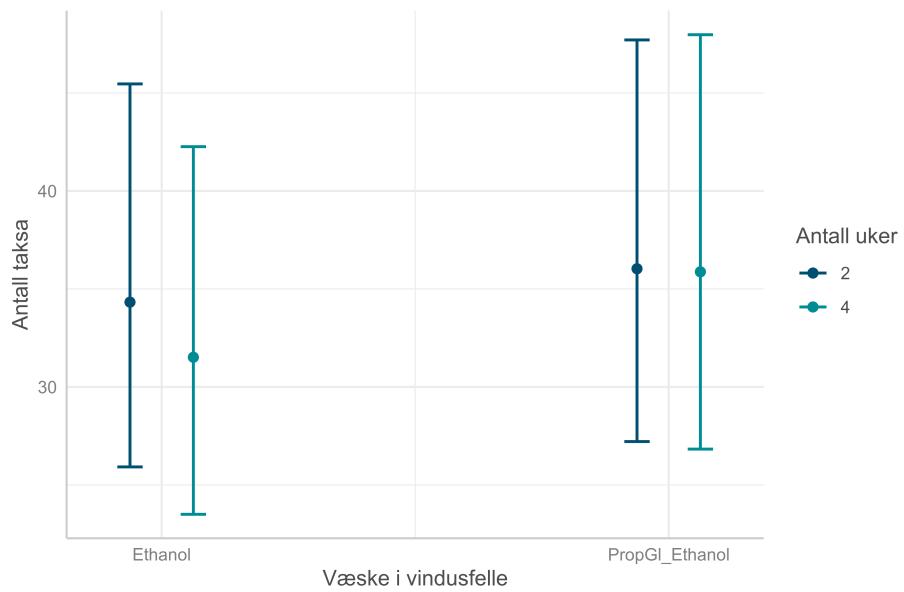


Figure 12: Modelled difference in number of taxa found in window traps, in 2/4 weeks sampling and ethanol/propylen-glycole liquid.